Vertex Sparsifiers: New Results from Old Techniques

Matthias Englert[*] University of Warwick

Anupam Gupta[†] Carnegie Mellon

Robert Krauthgamer[‡] Weizmann Institute

Harald Räcke[§] University of Warwick

Inbal Talgam-Cohen[‡] Weizmann Institute

Kunal Talwar[¶] MSR SVC

June 28, 2010

Abstract

Given a capacitated graph $G = (V, E)$ and a set of terminals $K \subseteq V$, how should we produce a graph H only on the terminals K so that every (multicommodity) flow between the terminals in G could be supported in H with low congestion, and vice versa? (Such a graph H is called a *flow-sparsifier* for G.) What if we want H to be a "simple" graph? What if we allow H to be a convex combination of simple graphs?

Improving on results of Moitra [FOCS 2009] and Leighton and Moitra [STOC 2010], we give efficient algorithms for constructing: (a) a flow-sparsifier H that maintains congestion up to a factor of $O(\frac{\log k}{\log\log k})$, where $k = |K|$. (b) a convex combination of trees over the terminals K that maintains congestion up to a factor of $O(\log k)$. (c) for a planar graph G, a convex combination of planar graphs that maintains congestion up to a constant factor. This requires us to give a new algorithm for the 0-extension problem, the first one in which the preimages of each terminal are connected in G. Moreover, this result extends to minor-closed families of graphs.

Our improved bounds immediately imply improved approximation guarantees for several terminal-based cut and ordering problems.

[*]Department of Computer Science and DIMAP, University of Warwick, Coventry, UK. Supported by EPSRC grant EP/F043333/1 and DIMAP (the Centre for Discrete Mathematics and its Applications).

[†]Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213. Research was partly supported by the NSF award CCF-0729022, and an Alfred P. Sloan Fellowship. This research was done when visiting Microsoft Research SVC, La Avenida, Mountain View CA.

[‡]This work was supported in part by The Israel Science Foundation (grant #452/08), and by a Minerva grant. Weizmann Institute of Science, Rehovot, Israel. Email: `{robert.krauthgamer,inbal.talgam@weizmann.ac.il}`

[§]Department of Computer Science and DIMAP, University of Warwick, Coventry, UK. Supported by DIMAP (the Centre for Discrete Mathematics and its Applications).

[¶]Microsoft Research Silicon Valley, 1065 La Avenida, Mountain View, CA, USA. `kunal@microsoft.com`.

1 Introduction

Given an undirected capacitated graph $G = (V, E)$ and a set of terminal nodes $K \subseteq V$, we consider the question of producing a graph H only on the terminals K so that the congestion incurred on G and H for any multicommodity flow routed between terminal nodes is similar. Often, we will want the graph H to be structurally "simpler" than G as well. Such a graph H will be called a *flow-sparsifier* for G; the *loss* (also known as *quality*) of the flow-sparsifier is the factor by which the congestions in the graphs G and H differ. For instance, when $K = V$, the results of [Räc08] give a convex combination of trees H with a loss of $O(\log n)$. We call this a *tree-based flow-sparsifier*—it uses a convex combination of trees.[1] Here and throughout, $k = |K|$ denotes the number of terminals, and $n = |V|$ the size of the graph.

For the case where $K \neq V$, it was shown by Moitra [Moi09] and by Leighton and Moitra [LM10] that for every G and K, there exists a flow-sparsifier $H = (K, E_H)$ whose loss is $O(\frac{\log k}{\log\log k})$, and moreover, one can efficiently find an $H' = (K, E_{H'})$ whose loss is $O(\frac{\log^2 k}{\log\log k})$. They used these to give approximation algorithms for several terminal-based problems, where the approximation factor depended poly-logarithmically on the number of terminals k, and not on n. We note that they construct an arbitrary graph on K, and do not attempt to directly obtain "simple" graphs; e.g., to get tree-based flow-sparsifiers on K, they apply [Räc08] to H', and increase the loss by an $O(\log k)$ factor.

In this paper, we simplify and unify some of these results: we show that using the general framework of interchanging distance-preserving mappings and capacity-preserving mappings from [Räc08] (which was reinterpreted in an abstract setting by Andersen and Feige [AF09]), we obtain the following improvements over the results of [Moi09, LM10].[2]

1. We show that using the 0-extension results [CKR04, FHRT03] in the framework of [Räc08, AF09] almost immediately gives us *efficent* constructions of flow-sparsifiers with loss $O(\frac{\log k}{\log\log k})$. While the existential result of [LM10] also used the connection between 0-extensions and flow sparsifiers, the algorithmically-efficient version of the result was done *ab initio*, increasing the loss by another $O(\log k)$ factor. We use existing machinery, thereby simplifying the exposition somewhat, and avoiding the increased loss.

2. We then use a randomized tree-embedding due to [GNR10], which is a variant of the so-called FRT tree-embedding [FRT04] where the expected stretch is reduced to $O(\log k)$ by requiring the non-contraction condition only for terminal pairs. Using this refined embedding in the framework of [Räc08, AF09], we obtain efficient constructions of *tree-based flow-sparsifiers* with loss $O(\log k)$.

3. We then turn to special families of graphs. For planar graphs, we give a new 0-extension algorithm that outputs a convex combination of 0-extensions $f : V \to K$ (with $f(x) = x$ for all $x \in K$), such that all the corresponding 0-extension graphs $H_f = (K, E_f)$ (namely, $E_f = \{(f(u), f(v)) : (u, v) \in E\}$) are *planar graphs*, and its expected stretch $\max_{u,v \in V} \mathbb{E}[d_{H_f}(f(u), f(v))]/d_G(u,v) \leq O(1)$. In particular, the planar graphs H_f produced are graph-theoretic minors of G. We remark that the known 0-extension algorithms [CKR04, AFH$^+$04, LN05] do not ensure planarity of H_f.

   It follows that planar graphs admit a *planar-based flow-sparsifier* (i.e., which is a convex combination of capacitated planar graphs on vertex-set K) with loss $O(1)$, and that we can find these efficiently. The fact that flow-sparsifiers with this loss *exist* was shown by [LM10], but their sparsifiers are not planar-based.

   Moreover, the 0-extension algorithm itself can be viewed as a randomized version of Steiner point removal in metrics: previously, it was only known how to remove Steiner points from tree metrics with $O(1)$ distortion [Gup01]. We believe this randomized procedure is of independent interest; e.g., combined with an

[1]Given a class $\mathcal{F}$ of graphs, we define an $\mathcal{F}$-flow-sparsifier to be a sparsifier that uses a single graph from $\mathcal{F}$ and an $\mathcal{F}$-based flow sparsifier to be a sparsifier that uses a convex combination of graphs from $\mathcal{F}$.

[2]Recently, it has come to our attention that, independent of and concurrent to our work, Charikar, Leighton, Li, and Moitra, and independently Makarychev and Makarychev, obtained results similar to the first two below, as well as related lower bounds.

embedding of [GNRS04], this gives an alternate proof of the fact that the metric induced on the vertices of a single face of a planar graph can be embedded into a distribution over trees [LS09].

4. The results for planar graphs are in fact much more general. Suppose G is a β_G-decomposable graph (see definition in Section 1.1). Then we can efficiently output a distribution over graphs $H_f = (K, E_f)$ such that these are all minors of G, and the expected stretch $\max_{u,v \in V} \mathbb{E}[d_{H_f}(f(u), f(v))]/d_G(u,v)$ is bounded by $O(\beta_G \log \beta_G)$. Now applying the same ideas of interchanging distance and capacity preservation, given any G and K, we can find *minor-based flow sparsifiers* with loss $O(\beta_G \log \beta_G)$.
5. Finally, we show some lower bounds on flow-sparsifiers: we show that flow-sparsifiers that are 0-extensions of the original graph must have loss at least $\Omega(\sqrt{\log k})$ in the worst-case. For this class of possible flow sparsifiers, this improves on the $\Omega(\log \log k)$ lower bound for sparsifiers proved in [LM10]. We also show that any flow-sparsifier that only uses edge capacities which are bounded from below by a constant, must suffer a loss of $\Omega(\sqrt{\log k}/\log \log k)$ in the worst-case.

We can use these results to improve the approximation ratios of several application problems. In many cases, constructions based on trees allow us to use better algorithms. Our results are summarized in Table 1. Note that apart from the two linear-arrangement problems, our results smoothly approach the best known results for the case $k = n$.

	Previous Best Result	Our Result	Best Result when $k = n$
Flow Sparsifiers (efficient)	$O(\frac{\log^2 k}{\log \log k})$	$O(\frac{\log k}{\log \log k})$	—
Tree-Based Flow Sparsifiers	$O(\log n)^\dagger$, $O(\frac{\log^3 k}{\log \log k})$	$O(\log k)$	$\Theta(\log n)$
Minor-based Flow Sparsifiers	—	$O(\beta_G \log \beta_G)$	—
Steiner Oblivious Routing	$\widetilde{O}(\log^2 k)$	$O(\log k)$	$\Theta(\log n)$
ℓ-Multicut	$\widetilde{O}(\log^3 k)$	$O(\log k)$	$O(\log n)$
Steiner Minimum Linear Arrangement (SMLA)	$\widetilde{O}(\log^{2.5} k)$	$O(\log k \log \log k)$	$O(\sqrt{\log n} \log \log n)$
SMLA in planar graphs	$\widetilde{O}(\log^{1.5} k)$	$O(\log \log k)$	$O(\log \log n)$
Steiner Min-Cut Linear Arrangement	$\widetilde{O}(\log^4 k)$	$O(\log^2 k)$	$O(\log^{1.5} n)$
Steiner Graph Bisection	$O(\log n)^\dagger$, $O(\frac{\log^3 k}{\log \log k})$	$O(\log k)$	$O(\log n)$

Table 1: Summary of our results. Previous results marked with † from [Räc08], all others from [Moi09, LM10].

Many of these applications further improve when the graph comes from a minor-closed family (and hence has good β-decompositions): e.g., for the Steiner Minimum Linear Arrangement problem on planar graphs, we can get an $O(\log \log k)$-approximation by using our minor-based flow-sparsifiers to reduce the problem to planar instances on the k terminals. Finally, in Appendix C we show how to get better approximations for the Steiner linear arrangement problems above using direct LP/SDP approaches.

1.1 Notation

Our graphs will have edge lengths or capacities; all edge-lengths will be denoted by $\ell : E \to \mathbb{R}_{\geq 0}$, and edge costs/capacities will be denoted by $c : E \to \mathbb{R}_{\geq 0}$. When we refer to a graph (G, ℓ), we mean a graph G with edge-lengths $\ell(\cdot)$; similarly (H, c) denotes one with capacities $c(\cdot)$. When there is potential for confusion, we will add subscripts (e.g., $c_H(\cdot)$ or $\ell_G(\cdot)$) for disambiguation. Given a graph (G, ℓ), the shortest-path distances under the edge lengths ℓ is denoted by $d_G : V \times V \to \mathbb{R}_{\geq 0}$.

Given a graph $G = (V, E)$ and a subset of vertices $K \subseteq V$ designated as *terminals*, a *retraction* is a map $f : V \to K$ such that $f(x) = x$ for all $x \in K$. For (G, c) and terminals $K \subseteq V$, a *K-flow in G* is a multicommodity flow whose sources and sinks lie in K.

Decomposition of Metrics. Let (X, d) be a metric space with terminals $K \subset X$. A partition (i.e., a set of disjoint "clusters") P of X is called *Δ-bounded* if every cluster $S \in P$ satisfies $max_{u,v \in S}\, d(u, v) \leq \Delta$. The metric (X, d) with terminals K is called *β-decomposable* if for every $\Delta > 0$ there is polynomial time algorithm to sample from a probability distribution μ over partitions of X, with the following properties:

- *Diameter bound:* Every partition $P \in \operatorname{supp}(\mu)$ is Δ-bounded.
- *Separation event:* For all $u, v \in X$, $\Pr_{P \in \mu}[\exists S \in P \text{ such that } u \in S \text{ but } v \notin S\,] \leq \beta \cdot d(u, v)/\Delta$.

β-decompositions of metrics have become standard tools with many applications; for more information see, e.g., [LN05].

We say that *a graph* $G = (V, E)$ is β-decomposable if for every nonnegative edge-lengths ℓ_G, the resulting shortest-path metric d_G is β-decomposable. Additionally, we assume that each cluster S in any partition P induces a *connected* subgraph of G; if not, break such a cluster into its connected components. The diameter bound and separation probabilities for edges remain unchanged by this operation; the separation probability for non-adjacent pairs (u, v) can be bounded by $\beta \cdot d(u, v)/\Delta$ by noting that some edge on the u-v shortest path must be separated for (u, v) to be separated, and applying the union bound.

2 0-Extensions

In this section we provide a definition of 0-extension which is somewhat different than the standard definition, and review some known results for 0-extensions. We also derive in Corollary 2.4 a variation of a known result on tree embeddings, which will be applied in Section 3.

A 0-*extension* of graph $(G = (V, E), \ell_G)$ with terminals $K \subseteq V$ is usually defined as a retraction $f : V \to K$. We define a 0-extension to be a retraction $f : V \to K$ along with another graph $(H = (K, E_H), \ell_H)$; here, the length function $\ell_H : E_H \to \mathbb{R}_+$ is defined as $\ell_H(x, y) = d_G(x, y)$ for every edge $(x, y) \in E_H$. Note that this immediately implies $d_H(x, y) \geq d_G(x, y)$ for all $x, y \in K$. Note also that H_f defined in Section 1 is a special case of H in which $E_H = \{(f(u), f(v)) : (u, v) \in E\}$, whereas, in general, H is allowed more flexibility (e.g., H can be a tree). This flexibility is precisely the reason we are interested both in the retraction f and in the graph H—we will often want H to be structurally simpler than G (just like we want a flow-sparsifier to be simpler than the original graph).

For a (randomized) algorithm $\mathcal{A}$ that takes as input (G, ℓ_G) and outputs a (random) 0-extension (H, ℓ_H), the *stretch factor* of algorithm $\mathcal{A}$ is the minimum $\alpha \geq 1$ such that

$$\mathbb{E}_H[\, d_H(f(x), f(y))\,] \leq \alpha\, d_G(x, y) \qquad \text{for all } x, y \in V.$$

The following are well-known results for 0-extension.

Theorem 2.1 ([FHRT03]) *There is an algorithm $\mathcal{A}_{FHRT}$ for 0-extension with stretch $\alpha = \alpha_{FHRT} := O(\frac{\log k}{\log\log k})$.*

Theorem 2.2 ([CKR04], see also [LN05]) *If the graph is β-decomposable, there is an algorithm $\mathcal{A}_{CKR}$ for 0-extensions with stretch $\alpha = \alpha_{CKR} := O(\beta)$.*

In particular, if the graph G belongs to a non-trivial family of graphs that is minor-closed, it follows from [KPR93, FT03] that $\alpha = O(1)$.

2.1 0-Extension With Trees

The following result is an extension of the tree-embedding theorem of Fakcharoenphol et al. [FRT04], where the difference is that the following result ensures the non-contracting property (a) only for terminal-terminal pairs, but replaces the $O(\log n)$ by $O(\log k)$ in the expected stretch between *any* pair of nodes.

Theorem 2.3 ([GNR10]) *There is a randomized polynomial-time algorithm that takes as input a graph $G = (V, E)$ with terminals $K \subseteq V$ and outputs a (random) edge-weighted 2-HST $T = (I \cup L, E_T)$ with internal nodes I and leaves L, and a map $f : V \to L$ such that*

(a) $d_T(f(x), f(y)) \geq d_G(x, y)$ *for all* $x, y \in K$ *(with probability 1),*
(b) $E_T[d_T(f(x), f(y))] \leq O(\log k)\ d_G(x, y)$ *for all* $x, y \in V$*, and*
(c) for each non-terminal $v \in V \setminus K$*, either there exists a terminal* x_v *sharing the leaf node with it (i.e.,* $f(v) = f(x_v)$*), or another descendent of* $f(v)$*'s parent in* T *contains a terminal* x_v*.*

Corollary 2.4 (Tree 0-extension) *There is a randomized polynomial-time algorithm* $\mathcal{A}_{GNR}$ *for 0-extension that has* $\alpha_{GNR} = O(\log k)$*; furthermore, the graphs output by the algorithm are trees on the vertex set* K*.*

Proof: To prove the corollary, we need to show an algorithm that takes as input a graph $G = (V, E)$ with terminals $K \subseteq V$ and outputs a (random) edge-weighted tree $T = (K, E)$ and a retraction $f : V \to K$ such that

(a') $d_T(x, y) \geq d_G(x, y)$ for all $x, y \in K$ (with probability 1),
(b') $E_T[d_T(f(x), f(y))] \leq O(\log k)\ d_G(x, y)$ for all $x, y \in V$.

We start with sampling a random tree $T' = (I \cup L, E')$ and associated map f from the distribution of Theorem 2.3. We can take any leaf $l \in L$ whose pre-image set only contains non-terminals, remove the leaf, and remap all $v \in f^{-1}(l)$ to some other leaf that is a descendent of l's parent node and also contains a terminal. (Such a leaf is guaranteed to exist by property *(c)* of Theorem 2.3.) While both the tree and the map change, we continue to call the modified tree T' and the map f. We repeat this process until all leaves in the modified tree T' contain at least one terminal. We can also assume that all terminals are at non-zero distance from each other (else we can remove some terminals, do the same proof, and add back in the terminals at the end)—now property *(a)* implies each leaf contains at most one terminal. Hence $f|_K$ is a 1-1 correspondence between the terminal set K and the remaining leaves in the tree T'. Since the tree T' is a 2-HST, the distances in the tree between a remapped non-terminal and any other node in T' (apart from the one it was identified with) do not change.

We can now remove all internal nodes in the modified version of T' (using, say, [Gup01]) to get a tree $T'' = (L, E'')$ on just the (erstwhile) leaves such that none of the $f(u)$-$f(v)$ distances are shrunk, and they are stretched by a factor of at most 8. The bijection between the set L and terminals K allows us to view the tree T'' as being on the node set K, and the map f as being a retraction from $V \to K$. Finally, shrinking the edges of the tree T'' only makes the expected stretch smaller, so we can reduce the length of any tree edge $e = (x, y)$ in T'' and set it equal to $d_G(x, y)$. Call this final tree T; it is immediate from properties *(a)* and *(b)* that this random T and the associated retraction $f : V \to K$ satisfy properties *(a')* and *(b')* above, where the big-Oh term in property (b') hides an extra stretch of 8 due to this post-processing. ■

As an aside, a weaker version of Corollary 2.4 with $O(\frac{\log^2 k}{\log\log k})$ can be proved as follows. First use Theorem 2.1 to obtain a random 0-extension H from G such that $\mathbb{E}_H[d_H(x, y)] \leq O(\frac{\log k}{\log\log k})\ d_G(x, y)$ for all $x, y \in K$. Then use the result of [FRT04] to get a random tree $H' = (K, E_{H'})$ such that $\mathbb{E}_{H'}[d_{H'}(x, y)] \leq O(\log k)\ d_H(x, y)$ for all $x, y \in V(H)$. Combining these two results proves the weaker claim.

3 Flow-Sparsifiers

Recall that given an edge-capacitated graph (G, c) and a set $K \subseteq V$ of terminals, a *flow-sparsifier with quality* ρ is another capacitated graph $(H = (K, E_H), c_H)$ such that (a) any feasible K-flow in G can be feasibly routed in

H, and (b) any feasible K-flow in H can be routed in G with congestion ρ.

3.1 Interchanging distance and capacity

We now use the framework of Räcke [Räc08], as interpreted by Andersen and Feige [AF09]. Given a graph $G = (V, E)$, let $\mathcal{P}$ be a collection of multisets of E, which will henceforth be called *paths*. A mapping $M : E \to \mathcal{P}$ maps each edge e to a path $M(e)$ in $\mathcal{P}$. Such a map can be represented as a matrix $\mathbf{M}$ in $\mathbb{Z}^{|E|\times|E|}$ where $\mathbf{M}_{e,e'}$ is the number of times the edge e' appears in the path (multiset) $M(e)$. Given a collection $\mathcal{M}$ of mappings (which we call the *admissible* mappings), a *probabilistic mapping* is a probability distribution over (or, convex combination of) admissible mappings; i.e., define $\lambda_M \geq 0$ for each $M \in \mathcal{M}$ such that $\sum_{M\in\mathcal{M}} \lambda_M = 1$.

Distance Mappings. Given $G = (V, E)$ and lengths $\ell : E \to \mathbb{R}_{>0}$,

- The *stretch* of an edge $e \in E$ under a mapping M is $\sum_{e'} \mathbf{M}_{e,e'}\ell(e')/\ell(e)$.
- The *average stretch* of e under a probabilistic mapping $\{\lambda\}$ is $\sum_M \lambda_M(\sum_{e'} \mathbf{M}_{e,e'}\ell(e')/\ell(e))$.
- The *stretch of a probabilistic mapping* is the maximum over all edges of their average stretch.

Capacity Mappings. Given a graph G with edge capacities $c : E \to \mathbb{R}_{>0}$,

- The *load* of an edge $e' \in E$ under a mapping M is $\sum_e \mathbf{M}_{e,e'}c(e)/c(e')$.
- The *expected load* of e' under a probabilistic mapping $\{\lambda\}$ is $\sum_M \lambda_M(\sum_e \mathbf{M}_{e,e'}c(e)/c(e'))$.
- The *congestion of a probabilistic mapping* is the maximum over all edges of their expected loads.

The Transfer Theorem. Andersen and Feige [AF09] distilled ideas from Räcke [Räc08] to state:

Theorem 3.1 (Theorem 6 in [AF09]) *Fix a graph G and a collection $\mathcal{M}$ of admissible mappings. For every $\rho \geq 1$, the following are equivalent:*

1. *For every collection of edge lengths ℓ_e, there is a probabilistic mapping with stretch at most ρ.*
2. *For every collection of edge capacities c_e, there is a probabilistic mapping with congestion at most ρ.*

In our settings, the techniques of Räcke [Räc08] can be used to make the result algorithmic: if one can efficiently sample from the probabilistic mapping with stretch ρ (which is true for the settings in this paper), one can efficiently sample from a probabilistic mapping with congestion $O(\rho)$ (and *vice versa*). In fact, one can obtain an explicit distribution on polynomially many admissible mappings. We defer further discussion of efficiency issues to the full version of the paper.

3.2 Tree-Based Flow Sparsifiers

The distance mappings we will consider will be similar to Räcke's application. Let us first fix for each $u, v \in K$ a canonical shortest-path S_{uv} between u, v in G. Now, consider a tree 0-extension (T, f) where $T = (K, E_T)$ and $f : V \to K$ is a retraction. For each edge $e = (w, x) \in E(G)$, consider the (unique) $f(w)$-$f(x)$-path $P_T(f(w), f(x))$ in the tree T. Define the mapping $M_T : E \to \mathcal{P}$ corresponding to the 0-extension (T, f) by

$$M_T((w, x)) = \uplus_{(u,v)\in P_T(f(w),f(x))} S_{uv}. \tag{3.1}$$

In other words, this maps each tree edge (w, x) to its canonical path; for each non-tree edge (w, x), it considers the edges on the tree-path between the images of w and x in the tree, and maps (w, x) to the disjoint union of the canonical paths of these edges. Recall that $M_T((w, x))$ is a multiset. In the corresponding matrix representation, $\mathbf{M}_{e,e'}$ is the multiplicity of e' in the set $\uplus_{(u,v)\in P_T(f(w),f(x))} S_{uv}$. Corollary 2.4 now implies the following:

Theorem 3.2 *Given a graph (G, ℓ) with terminals $K \subseteq V(G)$, there is a polynomial-time procedure to sample from a probabilistic mapping (which is a distribution over tree 0-extensions) with stretch $\rho_{dist} = O(\log k)$. Moreover, $\rho_{dist} \geq 1$ if $K \neq \emptyset$.*

Now we can apply the Transfer Theorem. Recall that in a K-flow, all source-sink pairs belong to set K.

Theorem 3.3 (Tree-Based Flow-Sparsifiers) *Given an edge-capacitated graph (G, c), and a set of terminals $K \subseteq V$, there is a polynomial-time algorithm that outputs a graph $H = (K, E_H)$ that is a convex combination of edge capacitated trees such that:*

(a) every K-flow that can be routed in G, can also be routed in H; and

(b) every K-flow that can be feasibly routed in H, can be routed with congestion $O(\log k)$ in G.

In other words, if we were to scale up the capacities in G to route all feasible flows in H, then the factor by which we would have to scale up capacities would only be $O(\log k)$.

Proof: We apply Theorem 3.1 and Theorem 3.2 to $G = (V, E)$ to get a convex combination $\{\lambda_{T,f}\}$ of maps $(T = (K, E_T), f)$ such that each edge in E has an average load of $O(\rho_{dist})$. Let us see how this implies (a) and (b) above: this is essentially a matter of unraveling the definitions. For each such (T, f), we define capacities on the edges $e_T \in E_T$ thus: let (A, B) be node sets of the two connected components of T formed by deleting the edge e_T, where $A \cup B = K$. Let $A' = \{v \in V \mid f(v) \in A\}$, and $B' = V \setminus A'$. Define

$$c_{T,f}(e_T) := \sum_{e \in E \cap (A' \times B')} c(e). \tag{3.2}$$

We claim that this convex combination $\{\lambda_{T,f}\}$ of capacitated trees satisfies (a) and (b). For (a), the definition of the capacities $c_{T,f}$ ensures that each edge of G can be concurrently routed feasibly in each T using capacities $c_{T,f}(\cdot)$, hence so can any K-flow feasible in G. Since this holds for each (T, f) pair, it holds for the convex combination.

To prove (b), we want to route edges in the convex combination of trees in the graph G, where we scale the capacities $c_{T,f}$ of edges from (T, f) by its convex multiplier $\lambda_{T,f}$. Consider any edge $e_T = (u, v) \in E_T$ with capacity $c_{T,f}(e_T)$ defined in (3.2): we can use the canonical shortest path S_{uv} to route this flow. Hence the load on any edge $e' = (w', x') \in E$ due to the convex combination of trees is at most

$$\frac{1}{c(e')} \sum_{T,f} \lambda_{T,f} \sum_{e_T \in E_T : e' \in S_{uv}} c_{T,f}(e_T). \tag{3.3}$$

Since $c_{T,f}(e_T)$ is the sum of the capacity of all edges $e = (w, x)$ such that e_T lies on the unique tree-path between $f(w)$, $f(x)$, we rewrite (3.3) as

$$\frac{1}{c(e')} \sum_{T,f} \lambda_{T,f} \sum_{e_T = (u,v) \in E_T : e' \in S_{uv}} \; \sum_{(w,x) \in E : e_T \in P_T(f(w), f(x))} c(wx) \tag{3.4}$$

$$= \frac{1}{c(e')} \sum_{T,f} \lambda_{T,f} \sum_{(w,x) \in E} c(wx) \times (\text{multiplicity of } e' \text{ in } \uplus_{(u,v) \in P_T(f(w), f(x))} S_{uv}). \tag{3.5}$$

However, this is exactly the *expected load* for e' under the notion of admissible maps defined in (3.1); hence this is bounded by the congestion (the maximum expected load over all edges), which is at most ρ_{dist} by Theorem 3.1. This proves condition (b) above, that the congestion to route any K-flow in the convex combination H in the graph G is at most ρ_{dist}. ■

3.3 General Flow Sparsifiers

Theorem 3.4 (Flow-Sparsifiers) *Given any graph G and terminals K, there is a randomized polynomial-time algorithm to output a flow-sparsifier H with loss $O(\frac{\log k}{\log\log k})$.*

Proof: Suppose we use Theorem 2.1 instead of using the tree 0-extension result (Corollary 2.4), we use the constructive version of the Transfer Theorem to get a polynomial number of graphs $H_1, H_2, \ldots$ on the vertex set K such that a convex combination of these graphs is a flow-sparsifier for the original graph G where the load is $O(\frac{\log k}{\log\log k})$. We can then construct a single graph H by setting the capacity of an edge to be the appropriate weighted combination of capacities of those edges in H_i; all feasible K-flows in G can be routed in H, and all feasible K-flows in H can be routed in G with congestion $O(\frac{\log k}{\log\log k})$. ■

The same idea using 0-extension results for β-decomposable graphs (Theorem 2.2) gives us the following:

Theorem 3.5 (Flow-Sparsifiers for Minor-Closed Families) *For any graph G that is β-decomposable and any K, there is a randomized polynomial-time algorithm to construct a flow-sparsifier with loss $O(\beta)$.*

Note that the decomposability holds if G belongs to a non-trivial minor-closed-family $\mathcal{G}$ (e.g., if G is planar). However, Theorem 3.5 does not claim that the flow-sparsifier for G also belongs to the family $\mathcal{G}$; this is the question we resolve in the next section.

4 Connected 0-Extensions and Minor-Based Flow-Sparsifiers

The results in this section apply to β-decomposable graphs. A prominent example of such graphs are planar graphs, which (along with every family of graphs excluding a fixed minor) are $O(1)$-decomposable [KPR93, FT03]. Thus, Theorem 4.1, Corollary 3.3 and Theorem 4.3 below all apply to planar graphs (and more generally to excluded-minor graphs) with $\beta = O(1)$. We now state our results for β-decomposable graphs in general. In Section 4.2 we define a related notion called *terminal-decomposability*, and show analogous results for $\hat{\beta}$-terminal-decomposable graphs.

In what follows we use the definition of 0-extension from Section 2 with $H = H_f$, i.e., $E_H = \{(f(u), f(v)) : (u, v) \in E\}$, hence the 0-extension is completely defined by the retraction f. We say that a 0-extension f is *connected* if for every x, $f^{-1}(x)$ induces a connected component in G. Our main result shows that we get connected 0-extensions with stretch $O(\beta \log \beta)$ for β-decomposable metrics.

Theorem 4.1 (Connected 0-Extension) *There is a randomized polynomial-time algorithm that, given $(G = (V, E), \ell_G)$ with terminals K such that d_G is β-decomposable, produces a connected 0-extension $f : V \to K$ such that for all $u, v \in V$, we have*

$$\mathbb{E}[d_H(f(u), f(v))] \leq O(\beta \log \beta) \cdot d_G(u, v).$$

Note that if f is a connected 0-extension, the graph H_f is a minor of G. Applying Theorem 3.1 to interchange the distance preservation with capacity preservation, we get the following analogue of Theorem 3.3.

Corollary 4.2 (Minor-Based Flow-Sparsifiers) *For every β-decomposable graph $G = (V, E)$ with edge capacities c_G and a subset $K \subset V$ of k terminals, there is a minor-based flow-sparsifier with quality $O(\beta \log \beta)$. Moreover, a minor-based flow-sparsifier for G, c_G, K can be computed efficiently in randomized poly-time.*

Since planar graphs are $O(1)$-decomposable and since their minors are planar, by Corollary 4.2 they have an efficiently constructable planar-based flow-sparsifier with quality $O(1)$. By Theorem 4.1, they always have a connected 0-extension with stretch at most $O(1)$. An interesting consequence of the latter result is that given any planar graph (G, ℓ_G), and a set K of terminals, we can "remove" the non-terminals and get a related planar graph on K while preserving inter-terminal distances in expectation. This generalizes a result of Gupta [Gup01] who showed a similar result for trees. (Obviously, this extends to every family of graphs excluding a fixed minor.)

Theorem 4.3 (Steiner Points Removal) *There is a randomized polynomial-time algorithm that, given* $(G = (V, E), \ell_G)$ *and* K *such that* d_G *is* β*-decomposable, outputs minors* $H = (K, E_H)$ *of* G *such that* $1 \leq \frac{\mathbb{E}[d_H(x,y)]}{d_G(x,y)} \leq O(\beta \log \beta)$ *for all* $x, y \in K$.

Note that these results only give us an $O(\log n \log\log n)$-approximation for connected 0-extension on arbitrary graphs (or an $O(\log^2 k \log\log k)$-approximation using results of Section 4.2). We can improve that to $O(\log k)$; the details are in Section 4.3.

Theorem 4.4 (Connected CKR) *There is a randomized polynomial-time algorithm that on input* $(G = (V, E), \ell_G)$ *and* K*, produces a connected* 0*-extension* f *with stretch factor* $\mathbb{E}[d_H(f(u), f(v))] \leq O(\log k) \cdot d_G(u, v)$ *for all* $u, v \in V$.

Using the semi-metric relaxation for 0-extension, we get a connected 0-extension whose cost is at most $O(\log k)$ times the optimal (possibly disconnected) 0-extension. To our knowledge, this is the first approximation algorithm for connected 0-extension, and in fact shows that the gap between the optimum connected 0-extension and the optimum 0-extension is bounded by $O(\log k)$. The same is true with an $O(1)$ bound for planar graphs. We remark that the connected 0-extension problem is a special case of the connected metric labeling problem, which has recently received attention in the vision community [VKR08, NL09].

4.1 The Algorithm for Decomposable Metrics

We now give the algorithm behind Theorem 4.1. Assume that edge lengths ℓ_G are integral and scaled such that the shortest edge is of length 1. Let the diameter of the metric be at most 2^δ. For each vertex $v \in V$, define $A_v = \min_{x \in K} d_G(v, x)$ to be the distance to the closest terminal. The algorithm maintains a partial mapping f at each point in time—some of the $f(v)$'s may be undefined (denoted by $f(v) = \perp$) during the run, but f is a well-defined 0-extension when the algorithm terminates. We say a vertex $v \in V$ is *mapped* if $f(v) \neq \perp$. The algorithm appears as Algorithm 1.

Algorithm 1 Algorithm for Connected 0-extension

```
 1: input: $(G, \ell_G), K$.
 2: let $i \leftarrow 0$, $f(x) = x$ for all $x \in K$, $f(v) = \perp$ for all $v \in V \setminus K$.
 3: while there is a $v$ such that $f(v) = \perp$ do
 4:     let $i \leftarrow i + 1$, $r_i \leftarrow 2^i$
 5:     sample a $\beta$-decomposition of $d_G$ with diameter bound $r_i$ to get a partition $P$
 6:     for all clusters $C_s$ in the partition $P$ that contains both mapped and unmapped vertices do
 7:         delete all vertices $u$ in $C_s$ with $f(u) \neq \perp$
 8:         for each connected component $C$ from $C_s$ do
 9:             choose a vertex $w_C \in C_s$ that was deleted and had an edge to $C$
10:             reset $f(u) = f(w_C)$ for all $u \in C$.
11:         end for
12:     end for
13: end while
```

We can assume that in round $\delta = \log \text{diam}(G)$, the partitioning algorithm returns a single cluster, in which case all vertices are mapped and the algorithm terminates. Let f_i be the mapping at the end of iteration i. For $x \in K$, let V_i^x denote $f_i^{-1}(x)$, the set of nodes colored x. The following claim follows inductively:

Lemma 4.5 *For every iteration* i *and* $x \in K$*, the set* V_i^x *induces a connected component in* G.

Proof: We prove the claim inductively. For $i = 0$, there is nothing to prove since $V_i^x = \{x\}$. Suppose that in iteration i, we map vertex u to x so that $u \in V_i^x$. Thus for some component C containing u, the mapped neighbor w_C chosen by the algorithm was in V_{i-1}^x. Since we map all of C to x, there is a path connecting v to w_C in V_i^x. Inductively, w_C is connected to x in $V_{i-1}^x \subseteq V_i^x$, and the claim follows. ■

The following lemma will be useful in the analysis of the stretch; it says that any node mapped in iteration i is mapped to a terminal at distance $O(2^i)$.

Lemma 4.6 *For every iteration i and $x \in K$, and every $u \in V_i^x$, $d_G(x, u) \le 2r_i$.*

Proof: The proof is inductive. For $i = 0$, the claim is immediate. Suppose that in iteration i, we map vertex u to x so that $u \in V_i^x$. Thus for some component C containing u, the mapped neighbor w_C chosen by the algorithm was in V_{i-1}^x. Moreover, u and w_C were in the same cluster in the decomposition so that $d(u, w_C) \le r_i$. Inductively, $d(w_C, x) \le 2r_{i-1}$ and the claim follows by triangle inequality. ■

In the rest of the section, we bound the stretch of the 0-extension; for every edge $e = (u, v)$ of G, we show that

$$\mathbf{E}[d_G(f(u), f(v)] \le O(\beta \log \beta)\ d_G(u, v).$$

Note that for $e = (u, v)$, $d_G((f(u), f(v)) = d_H((f(u), f(v))$, and so it's enough to prove the claim for d_G. The analogous claim for non-adjacent pairs will follow by triangle inequality, but here with d_H. We say that the edge $e = (u, v)$ is *settled in round j* if the later of its endpoints gets mapped in this round; e is *untouched after round j* if both u and v are unmapped at the end of round j. Let $d_G(u, K) \le d_G(v, K)$ and let A_e denote the distance $d_G(u, K)$. Let $j_e := \lfloor \log(A_e) \rfloor - 1$.

Lemma 4.7 *For edge $e = (u, v)$,*

(a) edge e is untouched after round $j_e - 1$,
(b) if edge e is settled in round j then $d_G(f(u), f(v)) = O(2^j + d_G(u, v))$.

Proof: For (a), if one of the end points of e is mapped before round j_e, then $2 \cdot 2^{j_e} \le A_e = d_G(e, K)$, which contradicts Lemma 4.6. For (b), both $d_G(u, f(u)), d_G(v, f(v)) \le 2^{j+1}$ by Lemma 4.6; the triangle inequality completes the proof. ■

Let $\mathcal{B}_j$ denote the "bad" event that the edge is settled in round j and that both end-points are mapped to different terminals. Let $z := \max\{A_e, d_G(u, v)\}$. We want to use

$$\mathbf{E}[d(f(u), f(v))] = \sum_j \mathbf{Pr}[\mathcal{B}_j] \cdot \mathbf{E}[d(f(u), f(v)) \mid \mathcal{B}_j].$$

Claim 4.8 $\mathbf{Pr}[\mathcal{B}_j] \le \min\{4\beta \frac{z}{2^j}, 1\} \cdot 5\beta \frac{d_G(u,v)}{2^j}$.

Proof: Recall that an edge is untouched after round j' if neither of its endpoints is mapped at the end of this round. For this to happen, u must be separated from its closest terminal in the clustering in round j', which happens with probability at most $\min\{\beta \frac{A_e}{2^{j'}}, 1\}$. Also recall that the probability that an edge $e = (u, v)$ is cut in a round j' is at most $\beta \frac{d_G(u,v)}{2^{j'}}$. Let i denote the round in which the edge is first touched. We upper bound the probability of the event $\mathcal{B}_j$ separately depending on how i and j compare. Note that for $j \le 2$, the right hand side is at least 1 so the claim holds trivially.

- $i \le j - 2$. For $\mathcal{B}_j$ to occur, the edge e must be cut in round $j - 2$ and $j - 1$, as otherwise it would already be settled in one of these rounds. The probability of this is at most $\min\{\beta \frac{d_G(u,v)}{2^{j-2}}, 1\} \cdot \beta \frac{d_G(u,v)}{2^{j-1}} \le \min\{4\beta \frac{z}{2^j}, 1\} \cdot 2\beta \frac{d_G(u,v)}{2^j}$.

- $i = j-1$. For $\mathcal{B}_j$ to occur, the edge e must be cut in round $j-1$ and must be untouched after round $j-2$. The probability of this is at most $\min\{\beta\frac{A_e}{2^{j-2}}, 1\} \cdot \beta\frac{d_G(u,v)}{2^{j-1}} \le \min\{4\beta\frac{z}{2^j}, 1\} \cdot 2\beta\frac{d_G(u,v)}{2^j}$.
- $i = j$. For $\mathcal{B}_j$ to occur, e must be cut in round j and must be untouched after round $j-1$. The probability of this is at most $\min\{\beta\frac{A_e}{2^{j-1}}, 1\} \cdot \beta\frac{d_G(u,v)}{2^j} \le \min\{4\beta\frac{z}{2^j}, 1\} \cdot \beta\frac{d_G(u,v)}{2^j}$.

Since $\mathbf{Pr}[\mathcal{B}_j] = \mathbf{Pr}[\mathcal{B}_j \wedge (i \le j-2)] + \mathbf{Pr}[\mathcal{B}_j \wedge (i = j-1)] + \mathbf{Pr}[\mathcal{B}_j \wedge (i = j)]$, the claim follows. ■

Lemma 4.7(b) implies that if the edge is settled before round $j_d := \lfloor \log(d_G(u,v)) \rfloor$, the conditional expectation $\mathbf{E}[d_G(f(u), f(v)) \mid \mathcal{B}_j]$ is $O(d_G(u,v))$. Moreover the edge e cannot be settled before round $j_e = \lfloor \log(A_e) \rfloor - 1$ by Lemma 4.7(a). Let $j_m := \max\{j_d, j_e\}$. It therefore suffices to to show that

$$\sum_{j \ge j_m} \mathbf{Pr}[\mathcal{B}_j] \cdot O(2^j) \le O(\beta \log \beta)\, d_G(u,v) \ .$$

Plugging in the upper bound for $\mathbf{Pr}[\mathcal{B}_j]$ into the left hand side, we get

$$\begin{aligned} \textstyle\sum_{j \ge j_m} \mathbf{Pr}[\mathcal{B}_j] \cdot O(2^j) &\le \textstyle\sum_{j \ge j_m} \min\{4\beta\frac{z}{2^j}, 1\} \cdot 5\beta\frac{d_G(u,v)}{2^j} \cdot O(2^j) \\ &\le \textstyle\sum_{j \ge j_m} \min\{4\beta\frac{z}{2^j}, 1\} \cdot \beta \cdot O(d_G(u,v)) \quad \le O(\beta \log \beta)\, d_G(u,v) \ . \end{aligned}$$

In the last step, we used that $z = \max\{A_e, d_G(u,v)\} \le \max\{2^{j_e+2}, 2^{j_d+1}\} \le 2^{j_m+2}$, so the first $O(\log \beta)$ terms contribute $O(\beta\, d_G(u,v))$, while the remaining terms form a geometric series and sum to $O(d_G(u,v))$. This completes the proof of Theorem 4.1.

4.2 Terminal Decompositions

The general theorem for connected 0-extensions gives a guarantee in terms of its decomposition parameter β, and in general this quantity may depend on n. This seems wasteful, since we decompose the entire metric while we mostly care about separating the terminals.

To this end, we define *terminal decompositions* (the reader might find it useful to contrast it with definition of decompositions in Section 1.1). A *partial partition* of a set X is a collection of disjoint subsets (called "clusters" of X). A metric (X, d) with terminals K is called $\hat{\beta}$*-terminal-decomposable* if for every $\Delta > 0$ there is probability distribution μ over partial partitions of X, with the following properties:

- *Diameter bound:* Every partial partition $\widehat{P} \in \mathrm{supp}(\mu)$ is connected and Δ-bounded.
- *Separation event:* For all $u, v \in X$, $\Pr_{\widehat{P} \in \mu}[\exists S \in \widehat{P}$ such that $u \in S$ but $v \notin S] \le \hat{\beta} \cdot d(u,v)/\Delta$.
- *Terminal partition:* For all $x \in K$, every partial partition $\widehat{P} \in \mathrm{supp}(\mu)$ has a cluster containing x.
- *Terminal-centered clusters:* For every partial partition $\widehat{P} \in \mathrm{supp}(\mu)$, every cluster $S \in \widehat{P}$ contains a terminal.

A graph $G = (V, E)$ with terminals K is $\hat{\beta}$-terminal-decomposable if for every nonnegative lengths ℓ_G assigned to its edges, the resulting shortest-path metric d_G with terminals K is $\hat{\beta}$-terminal-decomposable. Throughout, we assume that there is a polynomial time algorithm that, given the metric, terminals and Δ as input, samples a partial partition $\widehat{P} \in \mu$. Note that if $K = V$, the above definitions coincide with the definitions of β-decomposable metrics and graphs.

Our main theorem for terminal decomposable metrics is the following:

Theorem 4.9 *Given $(G = (V, E), \ell_G)$, suppose d_G is $\hat{\beta}$-terminal-decomposable with respect to terminals K. There is a randomized polynomial-time algorithm that produces a connected 0-extension $f : V \to K$ such that for all $u, v \in V$, we have $\mathbb{E}[d_G(f(u), f(v))] \le O(\hat{\beta}^2 \log \hat{\beta}) \cdot d_G(u,v)$.*

This theorem is interesting when $\hat{\beta}$ is much less than β, the decomposability of the metric itself. E.g., one can alter the CKR decomposition scheme to get $\hat{\beta}(k, n) = O(\log k)$, while $\beta = O(\log n)$.

4.2.1 The Modified Algorithm.

Algorithm 2 for the terminal-decomposable case is very similar to Algorithm 1: the main difference is that in each iteration we only obtain a partial partition of the vertices, we color only the nodes that lie in clusters of this partial partition.

A few words about the algorithm: recall that a partial partition returns a set of connected diameter-bounded clusters such that each cluster contains at least one terminal, and each terminal is in exactly one cluster— we use V^x to denote the cluster containing $x \in K$. (Hence either $V^x = V^y$ or $V^x \cap V^y = \emptyset$.) Now when we delete all the vertices in some cluster V^x that are already mapped, this includes the terminal x—and hence there is at least one candidate for w_C in Line 9. Eventually, there will be only one cluster, in which case all vertices are mapped and the algorithm terminates.

Algorithm 2 Algorithm for Connected 0-extension: the terminal-decomposable case

```
1:  input: $(G, \ell_G), K$.
2:  let $i \leftarrow 0$, $f(x) = x$ for all $x \in K$, $f(v) = \perp$ for all $v \in V \setminus K$.
3:  while there is a $v$ such that $f(v) = \perp$ do
4:      let $i \leftarrow i + 1$, $r_i \leftarrow 2^i$
5:      find a $\hat{\beta}$-terminal-decomposition of $d_G$ with diameter bound $r_i$; let $V^x$ be the cluster containing terminal $x$.
6:      for all clusters $V^x$ in the partial partition do
7:          delete all vertices $u$ in $V^x$ with $f(u) \neq \perp$
8:          for each connected component $C$ from $V^x$ thus formed do
9:              choose a vertex $w_C \in V^x$ that was deleted and had a neighbor in $C$
10:             reset $f(u) = f(w_C)$ for all $u \in C$.
11:         end for
12:     end for
13: end while
```

The analysis for Theorem 4.9 is almost the same as for Theorem 4.1; the only difference is that Claim 4.8 is replaced by the following weaker claim which immediately gives the $O(\hat{\beta}^2 \log \hat{\beta})$ bound.

Claim 4.10 $\mathbf{Pr}[\mathcal{B}_j] \leq \min\{8\hat{\beta}\frac{z}{2^j}, 1\} \cdot 23\hat{\beta}^2 \frac{d(u,v)}{2^j}$.

Proof: Recall that an edge is *untouched* after round j' if neither of its endpoints is mapped at the end of this round. For this to happen, u must be separated from it's closest terminal in the clustering in round j', which happens with probability at most $\min\{\hat{\beta}\frac{A_e}{2^{j'}}, 1\}$. Also recall that the probability that an edge $e = (u, v)$ is cut in a round j' is at most $\hat{\beta}\frac{d(u,v)}{2^{j'}}$. Let i denote the round in which the edge is first touched. We upper bound the probability of the event $\mathcal{B}_j$ separately depending on how i and j compare. Note that for $j \leq 3$, the right hand side is at least 1 so the claim holds trivially.

- $i \leq j - 3$. For $\mathcal{B}_j$ to occur, it must happen that the edge is cut in round i and it is either untouched or cut in rounds $j-1$ and $j-2$. The probability for this to happen is at most $\min\{\hat{\beta}\frac{d(u,v)}{2^i}, 1\} \cdot \min\{\hat{\beta}(\frac{A_e}{2^{j-2}} + \frac{d(u,v)}{2^{j-2}}), 1\} \cdot \hat{\beta}(\frac{A_e}{2^{j-1}} + \frac{d(u,v)}{2^{j-1}}) \leq \min\{\frac{d(u,v)}{2^i}, 1\}\min\{8\hat{\beta}\frac{z}{2^j}, 1\} \cdot 4\hat{\beta}^2\frac{z}{2^j}$. If $d(u, v) \geq A_e$ this is at most $\min\{8\hat{\beta}\frac{z}{2^j}, 1\} \cdot 16\hat{\beta}^2\frac{d(u,v)}{2^j}$ as $z = d(u, v)$. Otherwise, observe that $i \geq j_e$ as the edge cannot be touched before. Hence $2^i \geq A_e/4$, and plugging this in gives a bound of $\min\{8\hat{\beta}\frac{z}{2^j}, 1\} \cdot 16\hat{\beta}^2\frac{d(u,v)}{2^j}$, as well.

- $i = j - 2$. For $\mathcal{B}_j$ to occur, the edge e must be cut in round $j - 2$ and it must be cut or untouched in round $j - 1$, as otherwise it would already be settled in one of these rounds. The probability of this is at most $\hat{\beta}\frac{d(u,v)}{2^{j-2}} \cdot \min\{\hat{\beta}(\frac{d(u,v)}{2^{j-1}} + \frac{A_e}{2^{j-1}}), 1\} \leq \min\{4\hat{\beta}\frac{z}{2^j}, 1\} \cdot 4\hat{\beta}\frac{d(u,v)}{2^j}$.

- $i = j-1$. For $\mathcal{B}_j$ to occur, the edge e must be cut in round $j-1$ and must be untouched in round $j-2$. The probability of this is at most $\min\{\hat{\beta}\frac{A_e}{2^{j-2}}, 1\} \cdot \hat{\beta}\frac{d(u,v)}{2^{j-1}} \le \min\{4\hat{\beta}\frac{z}{2^j}, 1\} \cdot 2\hat{\beta}\frac{d(u,v)}{2^j}$.
- $i = j$. For $\mathcal{B}_j$ to occur, e must be cut in round j and must be untouched in round $j-1$. The probability of this is at most $\min\{\hat{\beta}\frac{A_e}{2^{j-1}}, 1\} \cdot \hat{\beta}\frac{d(u,v)}{2^j} \le \min\{4\hat{\beta}\frac{z}{2^j}, 1\} \cdot \hat{\beta}\frac{d(u,v)}{2^j}$.

Since $\mathbf{Pr}[\mathcal{B}_i] = \mathbf{Pr}[\mathcal{B}_i \wedge (i \le j-3)] + \mathbf{Pr}[\mathcal{B}_i \wedge (i = j-2)] + \mathbf{Pr}[\mathcal{B}_i \wedge (i = j-1)] + \mathbf{Pr}[\mathcal{B}_i \wedge (i = j)]$, the claim follows. ■

4.3 Connected 0-extension on General Graphs

Finally, we show that for general metrics, we can do better than the $O(\log^2 k \log\log k)$ guarantee implied by Theorem 4.9. In particular, we now prove Theorem 4.4, which gives a $O(\log k)$ guarantee. We still use Algorithm 1 from the previous section, but use a specific decomposition algorithm. The following result follows from Fakcharoenpol et al. [FHRT03], who built up on the work of Calinescu, Karloff and Rabani [CKR04]:

Theorem 4.11 ([FHRT03]) *Let $(G = (V, E), \ell_G)$ with a terminal set $K = \{x_1, \ldots, x_k\} \subseteq V$. There is a (randomized) polynomial-time algorithm that produces, for each $i = 0, 1, \ldots, \lceil \log \operatorname{diam}(G) \rceil$, a collection of $k+1$ clusters $\{C_0^i, C_1^i, \ldots, C_k^i\}$, such that*

(a) (Diameter) For any $j \neq 0$, C_j^i contains the terminal x_j, and $d(x_j, v) \le 2^i$ for any $v \in C_j^i$,
(b) (Separation) For any $u, v \in X$, $\Pr[\exists j$ such that $u \in C_j^i$ but $v \notin C_j^i] \le O(\beta_i^{uv}) \cdot d(u,v)/2^i$, where the probability is taken over the internal coin tosses of the algorithm, and
(c) (Amortization) For any $u, v \in X$, $\sum_i \beta_i^{uv} \le \beta = O(\log k)$.
(d) (Coverage) $\cup_{j\neq 0} C_j^i$ contains $\cup_{j=1}^k B_d(x_j, 2^{i-1})$.

We remark that we do not need each cluster to induce a connected component. Observe that the (Diameter) and (Coverage) properties imply:

(e) (Laminarity) For any i, $\cup_{j\neq 0} C_j^{i+1} \supseteq \cup_{j\neq 0} C_j^i$ with probability 1. Hence also $C_0^i \supseteq C_0^{i+1}$ with probability 1.

We run Algorithm 1 with this decomposition; the only worry is that since the clusters are not connected, it may be the case that in step 9, we may not find a node w_C as desired. In this case, we *expel* C from C_s, and do not map the vertices in C in this iteration. This ensures the connectivity property of $f_i^{-1}(x)$'s. Moreover, the Laminarity property inductively ensures that we never map any vertex from C_0^i by the end of round i. Since the diameter property bounds the diameter of every other cluster, Lemma 4.6 continues to hold.

Now, by its very definition, any expulsion operation only removes components that are disconnected from the rest of C_s, and hence does not increase the separation probability for any edge. Moreover, it is still the case the if u is mapped before round j and an edge (u, v) is not cut in round j, then the node v gets mapped in round j as well. Indeed by laminarity, u is in one of the clusters containing a terminal, and if (u, v) is not cut, then so is v. Since u is mapped, the component containing v cannot be expelled. Thus Claim 4.8 continues to hold and bounds the probability of $\mathcal{B}_j$, implying that

$$\begin{aligned}\mathbf{E}[d(f(u), f(v))] &= \textstyle\sum_j \mathbf{Pr}[\mathcal{B}_j] \cdot \mathbf{E}[d(f(u), f(v)) \mid \mathcal{B}_j] \\ &\le O(d_G(u,v)) + \textstyle\sum_{j\ge j'} \mathbf{Pr}[\mathcal{B}_j] \cdot O(2^j) \\ &\le O(d_G(u,v)) + \textstyle\sum_{j\ge j'} \min\{4\beta\frac{z}{2^j}, 1\} \cdot 5\beta_i^{uv}\frac{d_G(u,v)}{2^j} \cdot O(2^j) \quad \le O(\beta\, d_G(u,v)).\end{aligned}$$

Since $\beta = O(\log k)$, this gives us connected 0-extensions where the stretch is $O(\log k)$, and hence finishes the proof of Theorem 4.4.

5 Future Directions

We gave a set of results on and around the idea of flow-sparsifiers and 0-extensions. Some of these results are not tight, and it would be interesting to obtain better bounds for these problems. Another interesting direction for future work is this: define an ℓ*-sparse-extension* of graph $G = (V, E)$ with terminals K to be any graph $H = (Z, E_H)$ with $|Z| = \ell$, $K \subseteq Z \subseteq V$, along with a retraction $f : V \to Z$ that satisfies $d_H(x, y) \geq d_G(x, y)$ for all $x, y \in Z$. (Note that a $|K|$-sparse-extension is just a 0-extension; one possible $|V|$-sparse-extension is G itself.) What if we consider ℓ-sparse-extensions (H, f) with

$$E[\, d_H(f(x), f(y))\,] \leq \alpha\, d_G(x, y) \qquad \text{for all } x, y \in V,$$

where ideally $\ell = \mathrm{poly}(k)$, and $\alpha = O(1)$ (or just $\alpha \ll \frac{\log k}{\log \log k}$)? In other words, if we are willing to retain a small number of non-terminals, can we achieve better stretch bounds? Note that standard lower bounds for 0-extension have the property that $|V| = \mathrm{poly}(k)$—hence the entire graph G is a "good" solution ($\mathrm{poly}(k)$-sparse-extension with $\alpha = 1$).

References

[AF09] R. Andersen and U. Feige. Interchanging distance and capacity in probabilistic mappings. *CoRR*, abs/0907.3631, 2009.

[AFH+04] A. Archer, J. Fakcharoenphol, C. Harrelson, R. Krauthgamer, K. Talwar, and É. Tardos. Approximate classification via earthmover metrics. In *Proceedings of the 15th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1079–1087, New York, 2004. ACM.

[ARV09] S. Arora, S. Rao, and U. V. Vazirani. Expander flows, geometric embeddings and graph partitioning. *J. ACM*, 56(2), 2009.

[CKR04] G. Calinescu, H. J. Karloff, and Y. Rabani. Approximation algorithms for the 0-extension problem. *SIAM J. Comput.*, 34(2):358–372, 2004.

[ENRS00] G. Even, J. Naor, S. Rao, and B. Schieber. Divide-and-conquer approximation algorithms via spreading metrics. *J. ACM*, 47(4):585–616, 2000.

[FHRT03] J. Fakcharoenphol, C. Harrelson, S. Rao, and K. Talwar. An improved approximation algorithm for the 0-extension problem. In *Proceedings of the 14th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 257–265. Society for Industrial and Applied Mathematics, 2003.

[FRT04] J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. System Sci.*, 69(3):485–497, 2004.

[FT03] J. Fakcharoenphol and K. Talwar. Improved decompositions of graphs with forbidden minors. In *6th International workshop on Approximation algorithms for combinatorial optimization*, pages 36–46, 2003.

[GNR10] A. Gupta, V. Nagarajan, and R. Ravi. Improved approximation algorithms for requirement cut. *Operations Research Letters*, 38(4):322–325, 2010.

[GNRS04] A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Cuts, trees and ℓ_1-embeddings of graphs. *Combinatorica*, 24(2):233–269, 2004. (Preliminary version in 40th FOCS, 1999.).

[GNS06] D. Golovin, V. Nagarajan, and M. Singh. Approximating the K-multicut problem. In *Proceedings of the 17th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 621–630, 2006.

[Gup01] A. Gupta. Steiner points in tree metrics don't (really) help. In *Proceedings of the 12th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 220–227, 2001.

[HLW06] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the AMS*, 43(4):439–561, 2006.

[KPR93] P. Klein, S. A. Plotkin, and S. B. Rao. Excluded minors, network decomposition, and multicommodity flow. In *Proceedings of the 25th ACM Symposium on Theory of Computing (STOC)*, pages 682–690, 1993.

[LM10] T. Leighton and A. Moitra. Extensions and limits to vertex sparsification. In *Proceedings of the 42th ACM Symposium on Theory of Computing (STOC)*, pages 47–56, 2010.

[LN05] J. R. Lee and A. Naor. Extending Lipschitz functions via random metric partitions. *Invent. Math.*, 160(1):59–95, 2005.

[LR99] T. Leighton and S. B. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *J. ACM*, 46(6):787–832, 1999. (Preliminary version in *29th FOCS*, pages 422–431, 1988).

[LS09] J. R. Lee and A. Sidiropoulos. On the geometry of graphs with a forbidden minor. In *Proceedings of the 41st ACM Symposium on Theory of Computing (STOC)*, pages 245–254, 2009.

[Moi09] A. Moitra. Approximation algorithms for multicommodity-type problems with guarantees independent of the graph size. In *Proceedings of the 50th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 3–12, 2009.

[NL09] S. Nowozin and C. H. Lampert. Global connectivity potentials for random field models. In *Proceedings of the 22nd IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 818–825, 2009.

[Räc08] H. Räcke. Optimal hierarchical decompositions for congestion minimization in net works. In *Proceedings of the 40th ACM Symposium on Theory of Computing (STOC)*, pages 255–264, 2008.

[RR98] S. Rao and A. W. Richa. New approximation techniques for some ordering problems. In *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 211–218, 1998.

[VKR08] S. Vicente, V. Kolmogorov, and C. Rother. Graph cut based image segmentation with connectivity priors. In *Proceedings of the 21st IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

A Lower Bounds

In this section, we show two kinds of lower bounds. The first shows that any flow-sparsifier that is a convex combination of 0-extensions must suffer a loss of $\Omega(\sqrt{\log k})$—for such extension, this improves on the $\Omega(\log\log n)$ lower bound for (arbitrary) flow-sparsifiers [LM10]. The second shows that any flow-sparsifier that only uses edge capacities which are bounded from below by a constant, must suffer a loss of $\Omega(\sqrt{\log k}/\log\log k)$.

A.1 Lower Bounds for 0-extension-based Sparsifiers

The following result can be viewed as following from the duality between 0-extensions and 0-extension-based flow-sparsifiers (Theorem 3.1); by that theorem, not only do good 0-extension algorithms give good 0-extension-based flow-sparsifiers, the converse would also be true—and hence one can use a lower bound of Calinescu et al. [CKR04] to infer lower bounds on 0-extension-based flow-sparsifiers. The following theorem gives the explicit construction obtained thus.

Theorem A.1 *For infinitely many values of k, there is a graph $G = (V, E)$ and a set $K \subset V$ of size k for which any flow-sparsifier that is a convex combination of 0-extension graphs has quality at least $\Omega(\sqrt{\log k})$.*

Proof: We use the lower bound of $\Omega(\sqrt{\log k})$ on the 0-extension integrality ratio by Calinescu et al. [CKR04]. For completeness we describe their construction: Let G be an expander with n vertices, maximum degree Δ and expansion at least α, where Δ and α are fixed parameters. Define $l = \left\lceil \sqrt{\log n} \right\rceil$ and $k = \left\lceil \frac{n}{l} \right\rceil$. Choose any k distinct vertices $h_1, \ldots .h_k \in V(G)$ and add k new paths of length l starting at these vertices and ending at new vertices labeled $1, \ldots, k$. Denote the resulting graph by G' (note that $|V(G')| = O(n)$ and $|E(G')| = O(n)$), and let the terminals K be the new vertices $\{1, \ldots, k\}$. Set the capacities and lengths of the edges to 1. The distance $\mathrm{dist}_{G'}(s, t)$ between terminals s, t is set to be the shortest path distance in G' between s, t.

For the described instance G', K of the 0-extension problem, Calinescu et al. show a semimetric whose cost is $|E(G')|$, namely the shortest path metric $\mathrm{dist}_{G'}$ in G' with respect to edge lengths of 1. Indeed,

$$\textstyle\sum_{e=(u,v)\in E(G')} \mathrm{cap}_{G'}(e)\,\mathrm{dist}_{G'}(u,v) = |E(G')| = O(n).$$

On the other hand, they show that there exists a universal $c > 0$ such that for any 0-extension function $f : V(G') \to K$,

$$\textstyle\sum_{e=(u,v)\in E(G')} \mathrm{cap}_{G'}(e)\cdot \mathrm{dist}_{G'}(f(u), f(v)) \geq cn\sqrt{\log n} = \Omega\left(n\sqrt{\log k}\right).$$

We use the instance G', K to construct a feasible solution to the dual LP for which the objective value β is at least $\Omega\left(\sqrt{\log k}\right)$. To construct the feasible solution, we need to specify the values of the dual variables $\ell(e)$ for $e \in E(G')$, and $\mathrm{dist}(s,t)$ for $s,t \in K$. We set $\ell(e)$ to be $\frac{1}{|E(G')|}$. Thus, $\sum_{e\in E(G')} \mathrm{cap}_{G'}(e)\,\ell(e) = 1$. We set $\mathrm{dist}(s,t)$ to be the shortest path distance between terminals s,t in G' with respect to edge lengths $\ell(e)$. Clearly the specified values of $\ell(e), \mathrm{dist}(s,t)$ form a feasible solution to the dual LP for some value of β. By the above proof of Calinescu et al., for any 0-extension function $f : V(G') \to K$,

$$\textstyle\sum_{e=(u,v)\in E(G')} \mathrm{cap}_{G'}(e)\,\mathrm{dist}(f(u), f(v)) \geq \frac{cn\sqrt{\log n}}{|E(G')|} = \Omega\left(\sqrt{\log n}\right) = \Omega\left(\sqrt{\log k}\right).$$

Thus, we have show there exists a feasible solution to the dual LP such that $\beta > \Omega(\sqrt{\log k})$. This implies that for any convex combination of 0-extensions $H = \sum \lambda_i H_i$, the minimum congestion of routing $\vec{H}$ in G' is at least $\Omega(\sqrt{\log k})$, completing the proof. ■

A.2 Lower Bounds for Sparsifiers having no Small Edges

Theorem A.2 *For infinitely many values of k, there is a graph $G = (V, E)$ and a terminal set $K \subset V$ of size k for which any flow-sparsifier with edge capacities at least $\varepsilon > 0$ has quality at least $\Omega(\varepsilon\sqrt{\log k}/\log\log k)$.*

Proof: Let n be a sufficiently large prime. Let $G = (V, E)$ be a graph whose nodes correspond to the elements of $\mathbb{Z}_n$ and that contains an edge $\{u, v\}$ if $v = u + 1$, $v = u - 1$, or $v = u^{-1}$ (all operations are w.r.t. $\mathbb{Z}_n$ and we define 0^{-1} as 0.) In other words the graph consists of a Hamiltonian cycle plus some additional edges. This graph G is a 3-regular expander (see, e.g., [HLW06]).

Choose the set of terminals K as $\{i \cdot \lceil\sqrt{\log n}\rceil \mid 0 \leq i \leq k-1\}$, with $k = n/\lceil\sqrt{\log n}\rceil$. To simplify notation, we will omit floor- and ceiling-operations in the following. For $i \in [0, k-1]$, let B_i be the set of the $\sqrt{\log n}$ nodes on the Hamiltonian cycle between terminal i and $i+1$, including i but excluding $i+1$.

Let $H = (K, E_H)$ be a flow sparsifier for G with edge capacities at least $\varepsilon > 0$. Let d be the maximum weight degree of H, where the weighted degree of a node is the sum over all capacities of incident edges.

Claim A.3 *The maximum weighted degree d of H is at least*

$$c' \cdot \varepsilon \cdot \frac{\sqrt{\log n}}{\log\log n}$$

for some constant c'.

Proof: Consider a demand of $1/k$ between all pairs of terminals.

Since the minimum edge capacity is at least ε, the unweighted degree of H is at most d/ε. Due to this bounded degree, for sufficiently large k, there are at least $k^2/4$ terminal pairs that have distance at least $\log k/(2\log(d/\varepsilon))$ from each other (see e.g. [CKR04, Lemma 4.2]).

Each of these pairs induces a load of $1/k$ on at least $\log k/(2\log(d/\varepsilon))$ edges. Therefore, the total load in the network is at least $k\log k/(8\log(d/\varepsilon))$. Since H has at most $k \cdot d/(2\varepsilon)$ edges, the congestion in H is at least $\varepsilon \log k/(4d\log(d/\varepsilon))$.

The same demand can be routed with congestion at most $(c+1)\sqrt{\log n}$ in G, for some constant c depending on the edge expansion of G. Say each terminal i sends a total flow of 1. We can distribute this flow evenly between the nodes in B_i using only edges inside of B_i and with congestion of at most 1. This can easily be done, since we can send this flow along the Hamiltonian cycle to reach every node in B_i. Now, we route a uniform multicommodity flow on the whole expander, where the flow leaving each node is $1/\sqrt{\log n}$, i.e., the demand between every pair of nodes is $1/(n\sqrt{\log n})$. This requires congestion at most $c\log n \cdot (1/\sqrt{\log n}) = c\sqrt{\log n}$ [LR99]. Finally, the flow in each B_i is routed inside B_i to the respective terminal. Again, this can easily be done with congestion 1. In total, we sent a flow of $1/k$ between all pairs of terminals and the congestion is bounded by $c\sqrt{\log n}+2 \leq (c+1)\sqrt{\log n}$.

Hence, we identified a demand, that requires congestion at least $\varepsilon\log k/(4d\log(d/\varepsilon))$ in H but can be routed with congestion at most $(c+1)\sqrt{\log n}$ in G. Since H is a flow sparsifier, its congestion has to be bounded by the congestion in G and thus, $\varepsilon\log k/(4d\log(d/\varepsilon)) \leq (c+1)\sqrt{\log n}$. It follows that

$$\frac{d}{\varepsilon}\log\left(\frac{d}{\varepsilon}\right) \geq \frac{\log k}{4(c+1)\sqrt{\log n}} \quad .$$

Using the fact that $k = n/\sqrt{\log n}$, the claim follows. ■

Now pick a node in H that as weighted degree at least $c' \cdot \varepsilon \cdot \sqrt{\log n}/\log\log n$ (such a nodes exists due to Claim A.3). Consider the situation in which the demand between this node and every other node corresponds to the capacity of the edge connecting them in H, and all other demands are 0. Clearly, in H this can be routed with congestion 1. The terminal in G corresponding to node u, however, has only degree 3. Therefore, routing this demand in G results in congestion at least $c' \cdot \varepsilon \cdot \sqrt{\log n}/(3\log\log n) \geq c' \cdot \varepsilon \cdot \sqrt{\log k}/(3\log\log k)$, since that is the load on at least one of the outgoing edges of u. ■

B Applications

Most of these applications were considered by Moitra [Moi09], and Leighton and Moitra [LM10]; we show how our results above give improved approximations to the problems.

B.1 Steiner Oblivious Routing

Theorem 3.3 is an exact analogue of Räcke's theorem on general flows [Räc08] for the special case of K-flows, and hence immediately gives an $O(\log k)$-oblivious routing scheme for K-flows.

B.2 Steiner Minimum Linear Arrangement

Given $G = (V, E)$ and $K \subseteq V$ with $|K| = k$, the goal in the Steiner Minimum Linear Arrangement (SMLA) problem is to find a mapping $F : V \to [k]$ such that $F|_K : K \to [k]$ is a bijection. The goal is to minimize $\sum_{(u,v)\in E} c_{uv}|F(u) - F(v)|$. Note that for the non-Steiner MLA case where $K = V$, Rao and Richa [RR98] gave an $O(\log n)$-approximation for general graphs and an $O(\log\log n)$-approximation for graphs that admit $O(1)$-padded decompositions (which includes the family of all trees).

For our algorithm, we take a random tree/retraction pair (T, f) from the distribution above; this ensures that the cost of the optimal map F^* (viewed as a solution to the MLA problem on T) increases by an expected $O(\log k)$-factor. Now solving the MLA problem on the tree to within an $O(\log\log k)$ factor to get a map $\widehat{F_T} : K \to [k]$, and defining $\widehat{F}(x) = \widehat{F_T}(f(x))$ gives us an expected $O(\log k \log\log k)$-approximation. We show in Appendix C that this can be improved slightly to $O(\log k)$ using a more direct approach.

B.3 Requirement Cut

For requirement cut, [GNR10] already present an $O(\log k \log g)$-approximation.

B.4 Steiner Graph Bisection

In this problem, we are given a value k' and want to find a bipartition $(A, V \setminus A)$ of the graph such that $|A \cap K| = k'$, and that minimizes the cost of edges cut by the bipartition. The approach of Räcke, which embeds the graph G into a random tree and finds the best $(k', k - k')$ bipartition on that, gives us an $O(\log k)$ algorithm for this partitioning problem.

B.5 Steiner ℓ-Multicut

In this problem, we are given terminal pairs $\{s_i, t_i\}_{i \in [k]}$, and a value $k' \leq k$, and we want to find a minimum cost set of edges whose deletion separates at least k' terminal pairs. Again, we can embed the graph into a random tree losing an $O(\log k)$ factor, and use the theorem of Golovin et al. [GNS06] to get a $4/3 + \epsilon$-approximation on this tree; this gives us the randomized $O(\log k)$-approximation.

B.6 Steiner Min-Cut Linear Arrangement

The Steiner Min cut Linear Arrangement (SMCLA) problem is defined as follows: Given $G = (V, E)$ and $K \subseteq V$ with $|K| = k$, we want to find a mapping $F : V \to [k]$ such that $F|_K : K \to [k]$ is a bijection. The goal is to minimize $\max_i \sum_{x \in F^{-1}([i]), y \notin F^{-1}([i])} c_{xy}$. For the non-Steiner version of the problem, Leighton and Rao [LR99] show that given an α-approximation to the balanced partitioning (or to the bisection) problem, one can get an $O(\alpha \log n)$-approximation to the MCLA problem. Using [ARV09], this gives an $O(\log^{1.5} n)$-approximation to the MCLA problem.

We note that the reduction works immediately for the Steiner version of the problem: given an α-approximation to Steiner-bisection, one gets an $O(\alpha \log k)$-approximation to SMCLA. Thus we get an $O(\log^2 k)$-approximation to the SMCLA problem. We show in Appendix C that this can be improved to $O(\log^{1.5} k)$ using a more direct approach.

C Better Algorithms Using a Direct Approach

The vertex sparsifiers give a modular approach to solving steiner version of various problems. Not surprisingly, for some of these problems, a direct attack will lead to better algorithms. In this section, we show that applying known techniques for Minimum Linear Arrangement (MLA) problem lead to a better approximation ratio for Steiner MLA, and for Steiner Minimum Cut Linear Arrangement.

C.1 Steiner Minimum Linear Arrangement

Recall that the Steiner MLA problems is defined as follows. Given $G = (V, E)$ and $K \subseteq V$ with $|K| = k$, the goal is to find a mapping $F : V \to [k]$ such that $F|_K : K \to [k]$ is a bijection. The goal is to minimize $\sum_{(u,v) \in E} c_{uv} |F(u) - F(v)|$. Specifically, we show the following result:

Theorem C.1 *There is a polynomial time $O(\log k)$-approximation algorithm for the SMLA problem based on the natural linear programming relaxation.*

Proof: The linear program for the SMLA problem is based on the spreading metric linear programming relaxation for MLA introduced in [ENRS00].

$$\begin{array}{rrcll}
\text{min} & \sum_{(u,v)\in E} c_{uv} d_{uv} & & & \\
\text{subject to:} & & & & \\
\text{(Triangle Inequality)} & d_{uw} - d_{uv} - d_{vw} & \leq & 0 & \forall u, v, w \in V \\
\text{(Spreading)} & \sum_{v \in S} d_{uv} & \geq & \frac{|S|^2}{5} & \forall S \subseteq K, |S| \geq 2, u \in S \\
 & d_{uv} & \geq & 0 & \forall u, v \in V
\end{array}$$

It follows from [ENRS00] that the above is a valid linear programming relaxation to the SMLA problem, and that one can efficiently separate for the spreading constraints so that the LP can be solved in polynomial time using the Ellipsoid algorithm. Further, it is easy to check that the spreading constraints imply that for any $u \in K$, $|\mathbf{B}_d(u, r) \cap K| \leq 5r$. (Here, $\mathbf{B}_d(v, r) = \{w \mid d(v, w) \leq r\}$ is the "ball" around v of radius r in the metric d.)

Let d be a solution to the above linear program. Since d is a metric on V, it follows from Theorem 2.3 that we construct a (random) edge-weighted 2-HST $T = (I \cup K, E_T)$ with internal nodes I and leaves K, and a retraction $f : V \to K$ such that

(a) $d_T(f(x), f(y)) \geq d(x, y)$ for all $x, y \in K$ (with probability 1),
(b) $E_T[d_T(f(u), f(v))] \leq O(\log k)\, d(u, v)$ for all $u, v \in V$,

We argue that given this HST, we can construct a mapping $F_T : V \to [k]$ such that $F_T|_K : K \to [k]$ is a bijection. This mapping will have the property that $|F_T(u) - F_T(v)| \leq 5 d_T(f(u), f(v))$. The approximation ratio of $O(\log k)$ then follows from property (b) above.

The mapping F_T is defined by taking the natural left-to-right ordering on K defined by T, and assigning every other vertex $v \in V$ to the position $f(v)$. Formally, let π be a pre-order traversal of T. For every terminal $x \in K$, set $F_T(x)$ to the number of terminals in π that occur before x, i.e. $F_T(x) = |K \cap \{\pi_i : i \leq \pi^{-1}(x)\}|$. For every other vertex $u \in V$, set $F_T(u) = F_T(f(u))$. It is easy to check that $F_T|_K$ is a bijection.

We next upper bound $|F_T(u) - F_T(v)|$ for $u, v \in V$. Consider the terminals $t_u = f(u), t_v = f(v)$; if $t_u = t_v$, then $F_T(x) = F_T(y)$ and there is nothing to prove. Else let T_{uv} be the smallest subtree of T containing t_u and t_v. By the properties of the HST, we have $d_T(t_u, t_v) \geq d_T(t_u, z)$ for all $z \in T_{xy}$. Moreover, $d_T(u, v) = d_T(t_u, t_v)$. Now,

$$\begin{aligned}
|F_T(u) - F_T(v)| &= |F_T(t_u) - F_T(t_v)| \\
&\leq |K \cap T_{uv}| \\
&\leq |K \cap \mathbf{B}_{d_T}(t_u, d_T(t_u, t_v))| && \text{(Since } d_T(t_u, t_v) \geq d_T(t_u, z) \text{ for all } z \in T_{uv}) \\
&\leq |K \cap \mathbf{B}_d(t_u, d_T(t_u, t_v))| && \text{(By property (a))} \\
&\leq 5 d_T(t_u, t_v) && \text{(By the spreading property)} \\
&= 5 d_T(u, v).
\end{aligned}$$

This proves Theorem C.1 ■

C.2 Steiner Min Cut Linear Arrangement

Recall that the Steiner Min cut Linear Arrangement (SMCLA) problem is defined as follows. Given $G = (V, E)$ and $K \subseteq V$ with $|K| = k$, the goal is to find a mapping $F : V \to [k]$ such that $F|_K : K \to [k]$ is a bijection. The goal is to minimize $\max_i \sum_{x \in F^{-1}([i]), y \notin F^{-1}([i])} c_{xy}$. Specifically, we show the following result:

Theorem C.2 *There is a polynomial time $O(\log^{1.5} k)$-approximation algorithm for the SMCLA problem.*

The algorithm and the proof are the natural generalization of the $O(\log^{1.5} n)$ approximation to the min cut linear arrangement problem. We sketch the argument here.

This algorithm is based on an SDP formulation and the sparsest cut algorithm of [ARV09], who show the following

Theorem C.3 *There exist a constant $\varepsilon > 0$ such that the following holds. For any k-point ℓ_2^2 metric (S, d) satisfying $\sum_{x,y \in S} d_{xy} \geq \frac{|S|^2}{8}$, there are sets $A, B \subseteq S$ such that $|A|, |B| \geq \varepsilon k$ and $d(A, B) \geq \frac{\varepsilon}{\sqrt{\log k}}$. Moreover given vectors $\{v_x : x \in S\}$ representing d, such sets A, B can be found in polynomial time.*

Consider first the following linear program:

$$\begin{array}{rrcll} \text{min} & \sum_{(x,y)\in E} c_{xy} d_{xy} & & & \\ \text{subject to:} & & & & \\ \text{(Triangle Inequality)} & d_{xz} - d_{xy} - d_{yz} & \leq & 0 & \forall x, y, z \in V \\ \text{(Balance)} & \sum_{x,y \in K} d_{xy} & \geq & \frac{|K|^2}{8} & \\ & d_{xy} & \geq & 0 & \forall x, y \in V \end{array}$$

Let $F : V \to [k]$ be the optimum MCSLA with value OPT. Then the cut separating $F^{-1}([\lfloor \frac{k}{2} \rfloor])$ from its complement has value at most OPT, and gives a feasible integral solution to above linear program. Thus the value of the relaxation above is at most OPT.

Suppose in the above linear program, we additionally require that the distance metric d be an ℓ_2^2 metric, i.e. there exists vectors $v_x \in \Re^n$ such that $d(x, y) = ||v_x - v_y||_2^2$. This program can be naturally written as an SDP, and can be solved in polynomial time to return vectors $\{v_x\}$. Moreover, the optimum to this relaxation has value at most OPT as well. Theorem C.3 then implies that we can find sets $A, B \subseteq K$ such that $|A|, |B| \geq \varepsilon k$ and where $d(A, B) \geq \Delta = \frac{\varepsilon}{\sqrt{\log k}}$. Consider the sets $A_r = \{x \in V : d(A, x) \leq r\}$. For $0 < r < \Delta$, it is immediate that $A \subseteq A_r \subseteq V \setminus B$.

Picking r at random from $(0, \Delta)$, we observe that for any $x, y \in V$

$$\Pr[x \in A_r, y \notin A_r] \leq (d(y, A) - d(x, A))/\Delta,$$

so that by triangle inequality, the expected cost of the cut $(A_r, V \setminus A_r)$ is at most $\frac{1}{\Delta} \sum_{(x,y)\in E} c_{xy} d_{xy} \leq OPT/\Delta$. Thus we can find an $r \in (0, \Delta)$ such that

(a) $|K \cap A_r|, |K \cap (V \setminus A_r)| \leq (1 - \varepsilon)k$.
(b) $\sum_{x \in A_r, y \notin A_r} c_{xy} \leq O(OPT \sqrt{\log k})$.

We can recursively compute steiner linear arrangements for A_r and $V \setminus A_r$, and by condition (a), the depth of the recursion is at most $O(\log k)$. For any i, we can thus bound the total cost of edges from $F^{-1}([i])$ to $V \setminus F^{-1}([i])$. Indeed each level of the recursion contributes at most $O(OPT \sqrt{\log k})$ to this cost. Since there are at most $O(\log k)$ levels, we get an $O(\log^{1.5} k)$ approximation.